Multispectral Dynamic Codebook and Fusion Strategy for Moving Objects Detection

Rongrong Liu[(✉)], Yassine Ruichek, and Mohammed El Bagdouri

Connaissance et Intelligence Artificielle Distribuées (CIAD),
University Bourgogne Franche-Comté, UTBM, 90010 Belfort, France
{rongrong.liu,yassine.ruichek,mohammed.el-bagdouri}@utbm.fr

Abstract. The Codebook model is one of the popular real-time models for background subtraction to detect moving objects. In this paper, we propose two techniques to adapt the original Codebook algorithm to multispectral images: dynamic mechanism and fusion strategy. For each channel, only absolute spectral value is used to calculate the spectral similarity between the current frame pixel and reference average value in the matching process, which can simplify the matching equations. Besides, the deciding boundaries are obtained based on statistical information extracted from the data and always adjusting themselves to the scene changes. Results demonstrate that with the proposed techniques, we can acquire a comparable accuracy with other methods using the same multispectral dataset for background subtraction.

Keywords: Moving objects detection · Background subtraction · Multispectral images · Dynamic Codebook · Fusion strategy

1 Introduction

Moving object detection is one of the most commonly encountered low-level tasks in computer vision and a prerequisite of many intelligent video processing applications, such as automated video surveillance [1], object tracking [2,3] and parking lot management [4], to just name a few.

A common and widely used approach for detecting moving objects from stationary camera is background subtraction. As the name suggests, it is the process to automatically generate a binary mask which classifies the set of pixels into foreground and background, during which, the moving objects are always called the foreground and the static information is called the background. Thus, background subtraction is sometimes also known as foreground detection and foreground-background segmentation [5].

The last decade witnessed very significant publications on background subtraction. A quick search for "background subtraction" on IEEE Xplore returns over 2370 publications in the last ten years (2009–2019). Among these, Gaussian

Mixture Model (GMM) [6] modeled every pixel with a mixture of k Gaussian to handle multiple backgrounds. To deal with cluttered and fast variation scenes, a non-parametric technique known as Kernel Density Estimation (KDE) [7] was also proposed. There are also other methods that process images not only at the level of pixel, but also at region or even at frame levels [8]. Moreover, Kim et al. [9], with Codebook algorithm, summarized each background pixel by one or more codewords to cope with illumination changes. In our work, Codebook approach has been chosen as the base modeling frame for its simplicity and efficiency.

According to [10], most of the methods are based on color images, namely Red-Green-Blue (RGB). In recent years, thanks to the technological advances in video capture, multispectral imaging is becoming increasingly accepted and used by the computer vision and robotic communities, in particular, Benezeth et al. [11] present a publicly available collection of multispectral video sequences called multispectral video sequences (MVS) dataset. This is the first dataset on MVS available for research community in background subtraction.

In this paper, we try to utilize the different channels of multispectral images from a new perspective and propose a novel background subtraction framework motivated by Codebook algorithm. The goal is to simplify the similarity calculation between the current frame pixel and reference average value in the matching process.

The remainder of this paper is organized as follows. The proposed methods is presented in detail in Sect. 2. Section 3 demonstrates the experiments and comparisons with other methods on the MVS dataset. Final conclusions and future works are given in Sect. 4.

2   Proposed Multispectral Codebook Algorithms

The original Codebook background modeling algorithm was proposed in 2005 by Kim [9] to construct a background model from long observation sequences. It is quite simply a non-statistical clustering scheme with several important additional elements to make it robust against moving background. The motivation for adopting such a model is that it is fast to run, because it is deterministic; efficient for requiring little memory; adaptive; and able to handle complex backgrounds with sensitivity [12]. It has been proved to be very efficient in dealing with dynamic backgrounds.

During the last decade, many works have been dedicated to improve the original Codebook model [13,14]. For example, [15] has adopted multi-scale multi-feature codebook-based background subtraction targeting for challenging environments. In the work of [16], the background model is constructed by encoding each pixel into a codebook consisting of codewords based on a box model and it is also appropriate in the Hue-Saturation-Value (HSV) color space. Besides, [17] has rewritten the model parameters and then processed the three channels separately to simplify the matching equations. In [18], a dynamic boundary of codebook under the Lab color space has been developed.

2.1   Multispectral Dynamic Codebook

As a parametric method, the original Codebook needs parameter tuning to find the appropriate values for every scene. In this paper, we utilize a self-adaptive method motivated by [19], to select the optimal parameters automatically. In order of achieve this, some statistical information need to be calculated iteratively and recorded for each codeword during the whole process.

Multispectral Dynamic Codebook Model Initialization. At the beginning, the codebook is an empty set and the number of the codewords is set to 0. When the first multispectral frame comes, the Codebook model is initialized by constructing an associated codeword for each pixel. The corresponding vector $\mathbf{v}_m$ is set to be the average spectral values for all the channels and it is initialized as below:

$$\mathbf{v} = \mathbf{x} = (X_1, X_2, \ldots, X_n) \tag{1}$$

where n is the number of channels. The auxiliary information will be defined as a four-tuple $\mathbf{aux}_m = (f_m, \lambda_m, p_m, q_m)$, where f_m is the frequency of access or the number of times that the codeword is matched, λ_m is maximum length of time between consecutive accesses, p_m and q_m are the first and the last accesses times of the codeword respectively. They are initialized as:

$$\mathbf{aux} = (1, 0, 1, 1) \tag{2}$$

In spite of the vector $\mathbf{v}_m$ and the auxiliary tuple $\mathbf{aux}_m$, we also record a third vector named $\mathbf{S}_m$, which represents the set of the variance of the separate spectrum σ_i^2 and is initialized as:

$$\mathbf{S} = (X_1^2, X_2^2, \ldots, X_n^2) \tag{3}$$

What's more, another two vectors $\mathbf{B\_min}$ and $\mathbf{B\_max}$ need to be used to record the minimum and maximum values for each channel and they are initialized with the values of the spectral values of the first frame.

Multispectral Dynamic Codebook Matching Process. For an input pixel at time instant t, with the current value $\mathbf{x}_t = (X_1, X_2, \ldots, X_n)$, the matching codeword is found if

$$B\_low_i \leq X_i \leq B\_high_i \tag{4}$$

is satisfied for each channel, where B_low and B_high denote the lower and upper boundaries, respectively and are obtained by:

$$B\_low_i = max(B\_min_i - \sigma_i, \ 0) \tag{5}$$

$$B\_high_i = max(B\_max_i + \sigma_i, \ 255) \tag{6}$$

where σ_i is the standard deviation of the ith band value in the current codeword, whose square is the corresponding i^{th} element of $\mathbf{S}_m$. Thus, during this whole process, the boundaries are obtained from the data themselves and no manual parameters tuning is required, which makes the proposed model more practical.

Multispectral Dynamic Codebook Model Updating. When a new frame arrives, the matching criteria is first evaluated to see whether it is satisfied. If a match is found, the corresponding codeword is updated with the information of the current pixel, as illustrated in Algorithm 1, where N is the number of frames.

Algorithm 1. Multispectral dynamic Codebook construction

1: $L \leftarrow 0$, $\mathbf{C} \leftarrow \phi$
2: **for** $t = 1 \rightarrow N$ **do**
3:     $x_t = (X_1, X_1, ..., X_n)$
4:     Find the matching codeword to x_t in C if the following condition is satisfied for each channel.
5:     $B\_low_i \leq X_i \leq B\_high_i$
6:     **if** $\mathbf{C} = \phi$ or there is no match **then**
7:         $L \leftarrow L + 1$, create a new codeword $\mathbf{c}_L$
8:         $\mathbf{v}_0 = \mathbf{x}_t$
9:         $\mathbf{aux}_0 = (1, t - 1, t, t)$
10:        $\mathbf{S}_0 = (X_1^2, X_2^2, \ldots, X_n^2)$
11:        $\mathbf{B}\_min = (X_1, X_2, \ldots, X_n)$
12:        $\mathbf{B}\_max = (X_1, X_2, \ldots, X_n)$
13:    **else**
14:        update the matched codeword
15:        $\mathbf{v}_m \leftarrow (\frac{f_m \overline{X}_{m1} + X_1}{f_m + 1}, \frac{f_m \overline{X}_{m2} + X_2}{f_m + 1}, \ldots, \frac{f_m \overline{X}_{mn} + X_n}{f_m + 1})$
16:        $\mathbf{aux}_m \leftarrow (f_m + 1, max\lambda_m, t - q_m, p_m, t)$
17:        $\mathbf{S}_m \leftarrow (\frac{f_m \overline{\sigma}_{m1} + (X_1) - \overline{X}_{m1}^2}{f_m + 1}, \frac{f_m \overline{\sigma}_{m2} + (X_2) - \overline{X}_{m2}^2}{f_m + 1}, \ldots, \frac{f_m \overline{\sigma}_{mn} + (X_n) - \overline{X}_{mn}^2}{f_m + 1})$
18:        $B\_min_i \leftarrow min(B\_min_i, X_i)$
19:        $B\_max_i \leftarrow max(B\_max_i, X_i)$
20:    **end if**
21: **end for**

Multispectral Dynamic Codebook Model Foreground Detection. After constructing the Codebook model, the moving objects are detected by conducting the matching process. The pixel is detected as foreground if no acceptable matching codeword exists. Otherwise, it is classified as background and the corresponding codeword is updated according to the updating strategy.

2.2   Multispectral Fusion Strategy

Another step forward to exploit benefits of each spectral channel of multispectral images is to fuse the detection results of the monochromatic channels. The idea

is very straight forward. We first employ the multispectral dynamic Codebook which has been discussed in detail in the last subsection to every channel separately and obtain seven binary foreground-background masks independently. Then the detection results of the monochromatic channels (3 bands, 4 bands, 5 bands, 6 bands and 7 bands) are fused via union, vote or intersection to get the final foreground background segmentation result. The workflow for multispectral self-adaptive fusion strategy is shown in Fig. 1.
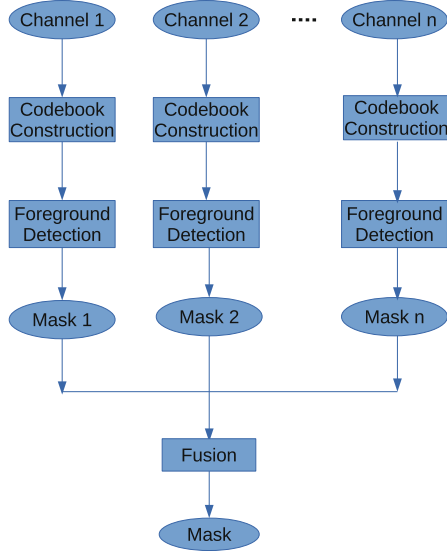

Fig. 1. Workflow of fusion strategy for multispectral self-adaptive Codebook

3   Experimental Results

To evaluate the performance of the proposed approaches for background subtraction, the MVS dataset presented by Benezeth et al. [11] is adopted for testing in this section. The proposed approaches are also compared with other methods using the same dataset [5]. Both visual and numerical results are displayed.

The MVS dataset contains a set of five challenging video sequences with seven multispectral channels or bands (six visible spectra and one near-infrared spectrum) captured simultaneously. These sequences are all publicly available, and the ground truth images are obtained manually. Note that the first scene is indoor, while the other four are outdoor.

The measure of accuracy employed in these experiments is F-measure, also known as balanced F-score or F1 score, which reaches its best value at 1 and worst score at 0. It is a harmonic mean of the precision and recall.

Firstly, the experiments for multispectral dynamic Codebook are conducted on the thirty-five different three-band-based combinations, thirty-five different four-band-based combinations, twenty-one different five-band-based combinations, seven different six-band-based combinations, and total seven-band, for the five videos composing the MVS dataset. Then the largest F-measures are selected and listed in Table 1. Accordingly, Table 2 shows the best F-measures with the fusion strategy. The largest F-measure for each video is in bold.

Table 1. Best F-measures with multispectral dynamic Codebook on the MVS dataset

Video	3 bands	4 bands	5 bands	6 bands	7 bands
1	0.8669	0.8750	**0.8757**	0.8749	0.7759
2	0.9607	**0.9609**	0.9607	0.9603	0.9307
3	0.9456	**0.9464**	0.9460	0.9445	0.8321
4	0.8693	**0.8731**	0.8734	0.8728	0.8661
5	0.9010	**0.9024**	0.8969	0.8966	0.8719

Table 2. Best F-measures with fusion strategy on the MVS dataset

Video	3 bands	4 bands	5 bands	6 bands	7 bands
1	**0.8967**	0.8836	0.8805	0.8712	0.7796
2	0.9607	**0.9608**	0.9607	0.9604	0.9575
3	**0.9372**	0.9346	0.9277	0.9020	0.8059
4	0.8685	0.8732	0.8741	**0.8742**	0.8709
5	0.8847	**0.8913**	0.8820	0.8646	0.8429

Figure 2 shows the visual results on the five video sequences. The top row is the original frames and the second row is the corresponding ground truth frames provided together with the dataset. The third and forth rows are the results obtained by the proposed multispectral dynamic Codebook and fusion strategy, respectively.

From Tables 1 and 2, the largest F-measure never appear when all seven channels are used. The combination with four channels has the largest possibility to achieve the best performance. This agrees with the assertion deduced with the Pooling method proposed by Yannick et al. [11] that only few channels actually define better the moving objects.

The results of the proposed methods are further compared in Table 3 with other methods [5] using the same dataset, in which, the brightness (B), spectral distortion (SD) and spectral information divergence (SID) are adopted to calculate the spectral distance in the matching process. The average F-measures for

Fig. 2. Visual results of the proposed methods on MVS dataset

all the five videos and outdoor sequences are calculated and listed in the last two rows.

From average F-measures for the whole dataset, our approaches produce comparable results to the best method that utilizes all the three criteria in [5] (4th column in Table 3), which proves the effectiveness of our methods. As we can see for the method adopting B and SD, the accuracy for outdoor scenes outperforms in average all the other mechanisms, but it achieves less satisfactory result for the indoor video. The two techniques proposed in this paper can be compromising solutions. Another advantage of the proposed methods is the low complexity of the matching equations, as the multispectral channels are processed separately only utilizing the intensity value of each channel and no correlation between channels need to be considered and calculated in the matching process. Besides, like other compared methods, it is also quite easy to adapt the proposed two algorithms to situation with any number of channels for background subtraction.

Table 3. F-measures with different methods on MVS dataset

Video	B+SD	B+SID	B+SD+SID	Proposed dynamic	Proposed fusion
1	0.8060	0.9219	0.9147	0.8757	0.8967
2	0.9643	0.9538	0.9642	0.9609	0.9608
3	0.9248	0.8939	0.9213	0.9464	0.9372
4	0.9001	0.8807	0.8979	0.8734	0.8742
5	0.9198	0.8842	0.8948	0.9024	0.8913
Mean	0.9030	0.9069	0.9186	0.9118	0.9121
Outdoor mean	0.9273	0.9032	0.9196	0.9208	0.9160

4   Conclusion and Future Works

In this paper, we have proposed two techniques to adapt the original Codebook algorithm to multispectral images: dynamic mechanism and fusion strategy, both of which process the seven channels of multispectral images independently. For each channel, only the intensity value is used to calculate the spectral similarity between the current frame pixel and reference one. Besides, the thresholds to determine are not set in advance empirically and fixed for the whole procedure, but obtained based on statistical information extracted from the data themselves and can always adjusting themselves to the scene changes. Results demonstrated that we can acquire a comparable accuracy using simpler matching equations than other techniques, when conducting experiments on the same multispectral public dataset. Our work may offer a new way for future works for applying multispectral images in moving objects detection.

Recently, deep learning-based methods have attracted huge attention in research community for its impressive performance for classification, semantic segmentation, localization, object detection and instance detection. Motivated by the recent success of deep neural networks for foreground segmentation [20], our next work is to explore deep learning besides traditional machine learning methods to investigate the benefits of multispectral images to improve the performance of background subtraction.

References

1. Khan, A., Janwe, N.J.: Review on moving object detection in video surveillance. Int. J. Adv. Res. Comput. Commun. Eng. **6**, 664–670 (2017)
2. Yang, T., Cappelle, C., Ruichek, Y., El Bagdouri, M.: Online multi-object tracking combining optical flow and compressive tracking in Markov decision process. J. Vis. Commun. Image Represent. **58**, 178–186 (2019)
3. Yang, T., Cappelle, C., Ruichek, Y., El Bagdouri, M.: Multi-object tracking with discriminant correlation filter based deep learning tracker. Integr. Comput.-Aided Eng. **26**, 1–12 (2019)
4. Yusnita, R., Norbaya, F., Basharuddin, N.: Intelligent parking space detection system based on image processing. Int. J. Innov. Manag. Technol. **3**(3), 232–235 (2012)
5. Liu, R., Ruichek, Y., El-Bagdouri, M.: Extended codebook with multispectral sequences for background subtraction. Sensors **19**(3), 703 (2019)
6. Bouwmans, T., El Baf, F., Vachon, B.: Background modeling using mixture of gaussians for foreground detection-a survey. Recent Pat. Comput.Sci. **1**(3), 219–237 (2008)
7. Elgammal, A., Harwood, D., Davis, L.: Non-parametric model for background subtraction. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 751–767. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45053-X_48
8. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: principles and practice of background maintenance. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, pp. 255–261. IEEE, Kerkyra (1999). https://doi.org/10.1109/ICCV.1999.791228

9. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. Real-Time Imaging **11**(3), 172–185 (2005)
10. Bouwmans, T.: Traditional and recent approaches in background modeling for foreground detection: an overview. Comput. Sci. Rev. **11**, 31–66 (2014)
11. Benezeth, Y., Sidibé, D., Thomas, J.B.: Background subtraction with multispectral video sequences. In: IEEE International Conference on Robotics and Automation Workshop on Non-Classical Cameras, Camera Networks and Omnidirectional Vision (OMNIVIS), Hong Kong, China (2014)
12. Doshi, A., Trivedi, M.: "Hybrid Cone-Cylinder" codebook model for foreground detection with shadow and highlight suppression. In: 2006 IEEE International Conference on Video and Signal Based Surveillance. IEEE, Sydney (2006). https://doi.org/10.1109/AVSS.2006.1
13. Liyun, G., Miao, Y., Timothy, G.: Online codebook modeling based background subtraction with a moving camera. In: 3rd International Conference on Frontiers of Signal Processing (ICFSP), pp. 136–140. IEEE, Paris (2017)
14. Mousse, M. A., Ezin, E.C., Motamed, C.: Foreground-background segmentation based on codebook and edge detector. In: Tenth International Conference on Signal-Image Technology and Internet-Based Systems, pp. 119–124. IEEE, Marrakech (2014)
15. Zaharescu, A., Jamieson, M.: Multi-scale multi-feature codebook-based background subtraction. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1753–1760. IEEE, Barcelona (2011). https://doi.org/10.1109/ICCVW.2011.6130461
16. Tu, Q., Xu, Y., Zhou, M.: Box-based codebook model for real-time objects detection. In: 2008 7th World Congress on Intelligent Control and Automation, pp. 7621–7625. IEEE, Chongqing (2008). https://doi.org/10.1109/WCICA.2008.4594112
17. Ruidong, G.: Moving object detection based on improved codebook model. In: 2nd International Conference on Modelling, Identification and Control. Atlantis, Paris (2015). https://doi.org/10.2991/mic-15.2015.4
18. Kusakunniran, W., Krungkaew, R.: Dynamic codebook for foreground segmentation in a video. Trans. Comput. Inf. Technol. (ECTI-CIT) **10**(2), 144–155 (2016)
19. Shah, M., Deng, J.D., Woodford, B.J.: A Self-adaptive CodeBook (SACB) model for real-time background subtraction. Image Vis. Comput. **38**, 52–64 (2015)
20. Lim, L.A., Keles, H.Y., Woodford, B.J.: Learning multi-scale features for foreground segmentation. Pattern Anal. Appl. **23**, 1369–1380 (2019). https://doi.org/10.1007/s10044-019-00845-9